Corrigé du Contrôle Continu n^o 1

Exercice 1

On a interrogé 100 vontants d'une circonscription et observé que 55% d'entre eux étaient favorable à un certain candidat.

1. Déterminons un intervalle centré en la proportion observée $\overline{f} = 0,55$ contenant p au seuil de 5%, c'est-à-dire avec probabilité 0,95. Puisque l'échantillon est de taille suffisante (ici $n = 100 \geq 30$), on sait que :

$$Z := \frac{\overline{f}_n - p}{\sqrt{\overline{f}_n(1-\overline{f}_n)/n}} \sim \mathcal{N}(0;1)$$

où $\overline{f}_n$ est la variable aléatoire (v.a.) représentant la proportion observée dans l'échantillon.
On doit déterminer un intervalle de la forme $I_{0,05} = [\overline{f} - l_{0,05}; \overline{f} + l_{0,05}]$ contenant p avec probabilité $0,95$. Pour cela, on écrit que :

$$\begin{aligned}
&\mathbf{P}\left[p \in I_{0,05}\right] = 0,95 \\
&\iff \mathbf{P}\left[\overline{f}_n - l_{0,05} \leq p \leq \overline{f}_n + l_{0,05}\right] = 0,95 \\
&\iff \mathbf{P}\left[\overline{f}_n - l_{0,05} \geq p \text{ ou } p \geq \overline{f}_n + l_{0,05}\right] = 0,05 \\
&\iff \mathbf{P}\left[\frac{\overline{f}_n - p}{\sqrt{\overline{f}_n(1-\overline{f}_n)/n}} \geq \sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}\, l_{0,05} \text{ ou } \frac{\overline{f}_n - p}{\sqrt{\overline{f}_n(1-\overline{f}_n)/n}} \leq -\sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}\, l_{0,05}\right] = 0,05 \\
&\iff \mathbf{P}\left[Z \geq \sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}\, l_{0,05} \text{ ou } Z \leq -\sqrt{\frac{n}{\overline{f}_n(1-\overline{f}_n)}}\, l_{0,05}\right] = 0,05.
\end{aligned}$$

Puisque Z suit la loi $\mathcal{N}(0;1)$, la table de la loi normale centrée réduite, dite « de l'écart-réduit », donne que

$$\mathbf{P}\left[Z \geq t_{0,05} \text{ ou } Z \leq -t_{0,05}\right] = 0,05$$

pour $t_{0,05} = 1,960$
Par identification, on obtient que :

$$\sqrt{\frac{n}{\overline{f}(1-\overline{f})}}\, l_{0,05} = t_{0,05} = 1,960,$$

soit encore

$$l_{0,05} = t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}} = 1,960\sqrt{\frac{0,55(1-0,55}{100}} \simeq 0,098.$$

Ainsi,

$$\begin{aligned}
I_{0,05} &= \left[\overline{f} - t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}}; \overline{f} + t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}}\right] \\
&\simeq [0,55-0,098; 0,55+0,098] = [0,452; 0,648].
\end{aligned}$$

2. On souhaite pouvoir affirmer, qu'avec la même fréquence observée, le candidat sera élu au risque d'erreur de 5% en ayant interrogé un nombre n suffisament grand de votants. Ceci revient à dire que l'on cherche pour quelles valeurs de n la borne inférieure de l'intervalle de confiance au seuil de 5% sera supérieure à 0,5. Pour cela, il faut et suffit que :

$$\overline{f} - t_{0,05}\sqrt{\frac{\overline{f}(1-\overline{f})}{n}} > 0,5$$

soit

$$0,55-1,960\sqrt{\frac{0,55(1-0,55)}{n}}>0,5$$

ou encore

$$1,960\sqrt{\frac{0,55(1-0,55)}{n}}<0,55-0,5=0,05$$

c'est-à-dire

$$1,960^2\frac{0,55(1-0,55)}{n}<0,05^2$$

c'est-à-dire

$$n>1,960^2\frac{0,55(1-0,55)}{0,05^2}\simeq 380,3.$$

Il faut interroger au moins 381 votants.

Exercice 2

Une expérience d'échantillonnage consiste à prélever 1000 échantillons de 64 bouteilles d'huile d'olive. La variable aléatoire "volume contenu réel" d'une bouteille est supposée normale de moyenne $50cl$ et d'écart-type $3cl$. On suppose les volumes contenus 2 à 2 indépendantes.

1. Déterminons la loi de la variable aléatoire $\overline{X}$: "volume contenu moyen dans un échantillon de 64 bouteilles". Notons X_i, $i=1,\ldots,50$, le volume contenu dans la i^{e} bouteille de l'échantillon. La variable aléatoire "volume contenu moyen dans une bouteille d'un échantillon de 64 bouteilles" s'écrit :

$$\overline{X}=\frac{X_1+\cdots+X_{64}}{64}=\frac{X_1}{64}+\cdots+\frac{X_{64}}{64}.$$

Puisque l'on sait que, pour tout i, $X_i\sim\mathcal{N}(50;3)$ et les X_i sont deux à deux indépendantes, on en déduit que $\overline{X}$ suit une loi normale de moyenne :

$$\mathbf{E}[\overline{X}]=\mathbf{E}\left[\frac{X_1}{64}+\cdots+\frac{X_{64}}{64}\right]=\frac{\mathbf{E}[X_1]}{64}+\cdots+\frac{\mathbf{E}[X_{64}]}{64}=\frac{50}{64}+\cdots+\frac{50}{64}=50$$

et de variance

$$\mathbf{V}[\overline{X}]=\mathbf{V}\left[\frac{X_1}{64}+\cdots+\frac{X_{64}}{64}\right]=\frac{\mathbf{V}[X_1]}{64^2}+\cdots+\frac{\mathbf{V}[X_{50}]}{64^2}=\frac{3^2}{64^2}+\cdots+\frac{3^2}{64^2}=\frac{9}{64}$$

c'est-à-dire d'écart-type

$$\sigma=\sqrt{\mathbf{V}[\overline{X}]}=\sqrt{\frac{9}{64}}=\frac{3}{8}.$$

Pour résumer :

$$\overline{X}\sim\mathcal{N}\left(50;\frac{3}{8}\right).$$

2. On a donc que :

$$Z=\frac{\overline{X}-50}{\frac{3}{8}}\sim\mathcal{N}(0;1).$$

La probabilité pour que la masse moyenne dans un échantillon soit inférieure à $49cl$ est :

$$\begin{aligned}\mathbf{P}\left[\overline{X}\le 49\right]&=\mathbf{P}\left[\frac{\overline{X}-50}{\frac{3}{8}}\le\frac{49-50}{\frac{3}{8}}\right]\\&=\mathbf{P}\left[Z\le-\frac{8}{3}\right]\\&=1-\mathbf{P}\left[Z\le\frac{8}{3}\right].\end{aligned}$$

Déterminons $\mathbf{P}\left[Z\le\frac{8}{3}\right]$ par interpolation linéaire. On a :

$$2,66\le\frac{8}{3}\le 2,67$$

donc, par croissance de la fonction de répartition et lecture dans la table de la fonction de répartition de la loi normale centrée et réduite,

$$0,9961 \simeq \mathbf{P}\left[Z \leq 2,66\right] \leq \mathbf{P}\left[Z \leq \frac{8}{3}\right] \leq \mathbf{P}\left[Z \leq 2,67\right] \simeq 0,9962.$$

On en déduit, par interpolation linéaire que :

$$\frac{\mathbf{P}\left[Z \leq \frac{8}{3}\right] - 0,9961}{0,9962 - 0,9961} \simeq \frac{\frac{8}{3} - 2,66}{2,67 - 2,66}$$

et donc

$$\mathbf{P}\left[Z \leq \frac{8}{3}\right] \simeq \frac{\frac{8}{3} - 2,66}{2,67 - 2,66}\left(0,9962 - 0,9961\right) + 0,9961 \simeq 0,9962.$$

Ainsi,

$$\mathbf{P}\left[\overline{X} \leq 49\right] = 1 - \mathbf{P}\left[Z \leq \frac{8}{3}\right] \simeq 1 - 0,9962 = 0,0038.$$

Parmi les 1000 échantillons, on peut s'attendre à ce que $1000 \times 0,0038 \simeq 3,8 \simeq 4$ présentent un volume contenu moyen inférieur à $49mg$.